# Tema 3: PROBABILIDAD - III

## Teorema de la probabilidad total. Teorema de Bayes. Tablas de contingencia y el lenguaje de las pruebas diagnósticas. Odds (probabilidades).

Biología sanitaria 2017/18. Universidad de Alcalá

M. Marvá. Actualizado: 2017-10-02

## La probabilidad total (sección 3.5.1 libro)

**Ejemplo** (obtenido de aquí): Si la prevalencia* de hipertensión en una cierta población es del 25%. Además, la prevalencia de infarto cardíaco
* Es del 0.3% para hipertensos. * Es del 0.1% para no hipertensos.

¿Cuál es la prevalencia del infarto en la población (total)?...¿de qué información dispones?

### Divide y vencerás

Cuando hay varios mecanismos/situaciones/escenarios... $B_1$, $B_2$,..., $B_k$ que pueden desencadenar un suceso $A$ y conocemos

- La probabilidad de que actúe cada uno de los mecanismos: $P(B_1),\ldots,P(B_k)$
- La probabilidad de que cada mecanismo $B_i$ desencadene $A$: $P(A|B_1),\ldots,P(A|B_k)$

Se puede calcular $P(A)$ mediante la estrategia de *divide y vencerás*

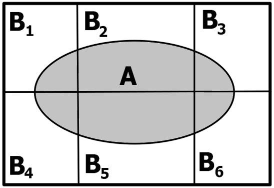
* **prevalencia**: proporción de individuos de un grupo o población que presentan una característica determinado

Hay que descomponer el espacio muestral en sucesos $B_1, \ldots, B_k$ tales que:

(1) $\Omega = B_1 \cup B_2 \cup \cdots \cup B_k$.

(2) $B_i \cap B_j = \emptyset$, para cualquier par $i \neq j$.

(3) $P(B_i) \neq 0$ para $i = 1, \ldots, k$.




Se dice que $B_1, \ldots, B_k$ es una partición (1) disjunta (2) del espacio muestral. Entonces:

**La probabilidad total (sección 3.3.2 libro)**

Si los sucesos $B_1, \ldots, B_K$ cumplen las condiciones (1), (2) y (3) entonces:

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\ &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_k)P(A|B_k) \end{aligned}$$

**Ejemplo** (continuación) (obtenido de aquí): Si la prevalencia de hipertensión en una cierta población es del 25% . Además, la prevalencia de infarto cardíaco

- Es del 0.3% para hipertensos.
- Es del 0.1% para no hipertensos.

¿Cuál es la prevalencia del infarto en esa población?

A = infarto, $B_1$=hipertenso, $B_2$=no hipertenso. Sabemos que

$$P(B_1) = 0.25, \qquad P(B_2) = 0.75, \qquad P(A|B_1) = 0.003, \qquad P(A|B_2) = 0.001$$

por tanto,

$$p(A) = 0.003x0.25 + 0.001x0.75 = 0.0015$$

## Conocido el efecto, ¿cuál es la probabiliadad de una causa?

**Ejemplo**: Si la prevalencia de hipertensión en una cierta población es del 25% . Además, la prevalencia de infarto cardíaco

- Es del 0.3% para hipertensos.
- Es del 0.1% para no hipertensos.

Si alguien ha tenido un infarto, ¿qué probabilidad hay de que no tuviera hipertensión?

$$P(B_1) = 0.25, \qquad P(B_2) = 0.75, \qquad P(A|B_1) = 0.003, \qquad P(A|B_2) = 0.001$$

por tanto,

$$p(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} = \frac{0.001x0.75}{0.003x0.25 + 0.001x0.75} = 0.5$$

# Conocido el efecto, ¿cuál es la probabiliadad de una causa?

**Ejemplo** (continuación) Si a una persona ha tenido infarto de miocardio, ¿qué probabilidad hay de que tenga hipertensión?

Es decir, una vez observado $A$, ¿qué probabilidad hay de que lo desencadenara $B_j$?

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}$$

y sabemos que

$$P(A|B_j) = \frac{P(B_j \cap A)}{P(B_j)}$$

de donde se deduce el

## Teorema de Bayes

Si los sucesos $B_1, \ldots, B_k$ cumplen las condiciones (1), (2) y (3) entonces:

$$P(B_j|A) = \frac{P(B_k)P(A|B_k)}{P(A)} = \frac{P(B_k)P(A|B_k)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_k)P(A|B_k)}$$

Se usa la terminología

- **Probabilidad a priori** para $P(A|B_j)$.
- **Probabilidad a posteriori** para $P(B_j|A)$.

# ORGANIZACIÓN DE LA INFORMACIÓN: árboles vs tablas

**Ejemplo**: se aplica un test diagnóstico rápido a 20 enfermos (da positivo en 18 ) y 30 sanos (da positivo en 5).

| | $D^+$ | $D^-$ | Total |
|---|---|---|---|
| $+$ | 18 | 5 | 23 |
| $-$ | 2 | 25 | 27 |
| Total | 20 | 30 | 50 |

dividir entre 50

| | $D^+$ | $D^-$ | Total |
|---|---|---|---|
| $+$ | 0.36 | 0.1 | 0.46 |
| $-$ | 0.04 | 0.5 | 0.54 |
| Total | 0.4 | 0.6 | 1 |

Conocemos la probabilidad de las intersecciones y de los eventos "marginales".

**Ejemplo**: la prevalencia de cierta enfermedad es del 40%. El 90% de los enfermos da positivo en test y 1 de cada 6 individuos sanos da positivo en el test.

$$\begin{cases} P(D^+) = 0.4 \begin{cases} P(+|D^+) = 0.9 \\ \\ P(-|D^+) = 0.1 \end{cases} \\ \\ P(D^-) = 0.6 \begin{cases} P(+|D^-) = 1/6 \\ \\ P(-|D^-) = 5/6 \end{cases} \end{cases}$$

Conocemos algunas probabilidades marginales y condicionadas

En el contexto de las pruebas diagnósticas

| | | **Enfermedad:** | | |
|---|---|---|---|---|
| | | Enfermos $D$ | Sanos $D^-$ | Total |
| **Resultado de la prueba:** | Positivo $+$ | $n_{11}$ | $n_{12}$ | $n_{1+}$ |
| | Negativo $-$ | $n_{21}$ | $n_{22}$ | $n_{2+}$ |
| | Total | $n_{+1}$ | $n_{+2}$ | $n$ |

- la **prevalencia de la enfermedad** es $P(D) = \dfrac{n_{+1}}{n}$
- El **valor predictivo positivo** de la prueba es

$$VPP = P(D \,|\, +) = \frac{n_{11}}{n_{1+}}.$$

- El **valor predictivo negativo** de la prueba es

$$VPN = P(D^- \,|\, -) = \frac{n_{22}}{n_{2+}}.$$

- Podemos llamar **valor predictivo** de la prueba

$$P((D \cap +) \cup (D^- \cap -)) = \frac{n_{11} + n_{22}}{n}.$$